Please replace the paragraph beginning at **page 25, line 28** with the following rewritten paragraph:

To eliminate some bin aligned orientations from consideration, those bin aligned orientations having one or more common atom pairs with centers in aligned bins are found. For each one of these pairs, an overlap value of one is added to a total overlap value for the two strings at this orientation, thus producing an integer overlap value corresponding to a count of the common atom pairs having their centers in aligned bins. Because two atoms may be overlapping, but have their centers in non-aligned bins, a strict upper bound for a given bin aligned orientation may be taken to be the sum of the integer values for all the bin aligned positions within plus or minus two bin offsets from the one computed. For the molecules of Figure 14, for example, the upper bound is taken to be seven rather than three for the zero offset bin aligned orientation. The method of combining the totals from nearby offsets to determine the strict upper bounds is shown in Figure 15. The cross-hatching of the boxes indicate the different atom types that contribute each unit of approximate overlap. Once these upper bounds are computed, actual estimates using the occupation numbers as described above are performed for the bin aligned offsets having the highest upper bounds first. As above, once a real estimate has been produced which is higher than all remaining upper bounds, the computation may terminate, and the highest estimate is used to determine the upper limit of total overlap for the two molecules.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

REMARKS

The specific changes to the specification and the amended claims are shown on a separate set of pages attached hereto and entitled **VERSION WITH MARKINGS TO SHOW**

CHANGES MADE, which follows the signature page of this Amendment. On this set of pages, the insertions are underlined while the ~~deletions are stricken through~~.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 3/25/02

By: _____

Thomas R. Arno
Registration No. 40,490
Attorney of Record
620 Newport Center Drive
Sixteenth Floor
Newport Beach, CA 92660
(619) 235-8550

S:\DOCS\ETA\ETA-1092.DOC
031402

VERSION WITH MARKINGS TO SHOW CHANGES MADE

Paragraph at **page 4, line 24:**

In one such embodiment, therefore, an error metric S^2, defined as follows:

~~(1)~~ $$S^2 = \frac{\sum_{i>j}^{n} \left(|x_i - x_j| - d_{ij} \right)^2}{\sum_{i>j}^{n} (d_{ij})^2}$$

Paragraph at **page 5, line 3:**

It may in some cases be advantageous to use alternative versions of S^2. For example, there are alternatives for this formula from the distance geometry literature which are also suitable for use in conjunction with the invention. Several of these can be found in "The Theory and Practice of Distance Geometry", T.F. Havel, I.D. Kuntz, and G.M. Crippen, Bull. Math. Biol., vol. 45, pp. 665-720 (1983), the entire disclosure of which is hereby incorporated by reference in its entirety. One such alternative function is:

~~1a)~~ $$S^2 = \sum_{i>j}^{n} \left[1 - \left(\frac{x_i - x_j}{d_{ij}} \right)^2 \right]^2$$

Paragraph at **page 5, line 13:**

Havel et al point out that this function exhibits good behavior for optimization purposes. Another possible function is:

~~1b)~~ $$S^2 = \sum_{i>j}^{n} \left\{ 1 - \left[\max(x_{ij} / d_{ij}, d_{ij} / x_{ij}) \right]^2 \right\}^2 .$$

Paragraph at **page 16, line 30:**

Once the maximum overlap value is determined, a molecular similarity score Sim_{AB} can be defined on the interval from 0 to 1 by normalizing the maximum overlap measured as follows:

$$Sim_{AB} = \frac{S_{\max,AB}}{\sqrt{S_{\max,AA}\ S_{\max,BB}}} \quad (3)$$

Paragraph at **page 22, line 30:**

To perform this comparison, string A and string B are oriented with their centers aligned. Then, the position of string B is shifted to align, as closely as possible, common atom pairs between the two strings. The amount of this shift Δx_B is calculated as follows:

$$\Delta x_B = \frac{1}{N_{match}} \sum_{\substack{Common \\ atom \\ pairs}} (x_A - x_B) \quad (4)$$

Paragraph at **page 23, line 7:**

After aligning the strings in this way, the squares of the linear offsets between all atom pairs of the same class in string A and string B is computed to produce a sum-squared-deviation (SSD) as follows:

$$SSD = \sum_{\substack{Common \\ atom \\ pairs}} (x_A - x_B)^2 \quad (5)$$

Paragraph at **page 23, line 21:**

Figures 11 and 12 illustrate the results of a comparison between a first compound, denoted compound A with two other compounds, denoted B1 and B2. Similarity calculations were performed using both 3D atomic coordinates to derive 1D representations, and also using 2D topological information to derive 1D representations. Figure 11 shows the result of the comparison between compound A and compound B1 when 3D and 2D information was used as a starting point. Figure 12 shows the result of the comparison between compound A and compound B2 when 3D

and 2D information was used as a starting point. Although graphs of overlap as a function of offset are shown in Figures 11 and 12 for illustrative purposes, it will be appreciated that in accordance with the above described techniques, most of the computations needed to generate such graphs are not required to be performed in order to produce the desired similarity measure. Using the equation: ~~equation 3 as set forth above~~

$$Sim_{AB} = \frac{S_{\max,AB}}{\sqrt{S_{\max,AA}\ S_{\max,BB}}}$$

the similarity value Sim_{AB1} for compounds A and B1 is 0.564, when 3D coordinates are used to derive the 1D representations, and is 0.529 when 2D topology is used to derive the 1D representations. In addition, the similarity value Sim_{AB2} for compounds A and B2 is 0.709, when 3D coordinates are used to derive the 1D representations, and is 0.775 when 2D topology is used to derive the 1D representations.

Paragraph at **page 25, line 23:**

Although this procedure is fast, ~~O~~one problem with ~~this procedure~~it is the fact that there are usually a large number of bin aligned orientations to consider. This number can be reduced in a manner analogous to that described above by computing upper bounds for each bin aligned position, and then eliminating from consideration those bin aligned orientations having upper bounds lower than a previously computed estimate. This is illustrated in Figures 14-15.

Paragraph at **page 25, line 28:**

To eliminate some bin aligned orientations from consideration, those bin aligned orientations having one or more common atom pairs with centers in aligned bins are found. For each one of these pairs, an overlap value of one is added to a total overlap value for the two strings at this orientation, thus producing an integer overlap value corresponding to a count of the common atom pairs having their centers in aligned bins. Because two atoms may be overlapping, but have their centers in non-aligned bins, a strict upper bound for a given bin aligned orientation may be taken to be the sum of the integer values for all the bin aligned positions within plus or minus two bin offsets from the one computed. For the molecules of Figure 14, for example, the upper bound is taken to be seven rather than three for the zero offset bin aligned orientation. The method of

combining the totals from nearby offsets to determine the strict upper bounds is shown in Figure 15. The cross-hatching of the boxes indicate the different atom types that contribute each unit of approximate overlap. Once these upper bounds are computed, actual estimates using the occupation numbers as described above are performed for the bin aligned offsets having the highest upper bounds first. As above, once a real estimate has been produced which is higher than all remaining upper bounds, the computation may terminate, and the highest estimate is used to determine the upper limit of total overlap for the two molecules.